

1. [Unit 5 Overview](#)
2. Probability
 1. [Why Probability Matters](#)
 2. [Probability](#)
 3. [The Meaning of "And" and "Or" in Probability](#)
 4. [Statistical Distributions](#)
 5. [Factorials](#)
 6. [Combinations](#)
3. Entropy
 1. [Introduction](#)
 2. [Gibbs Free Energy](#)
 3. [Second Law of Thermodynamics](#)
 4. [Example: Arranging energy and entropy](#)
 5. [A Way to Think about Entropy - Sharing](#)
 6. [Why Entropy is Logarithmic](#)

Unit 5 Overview

Exercise:



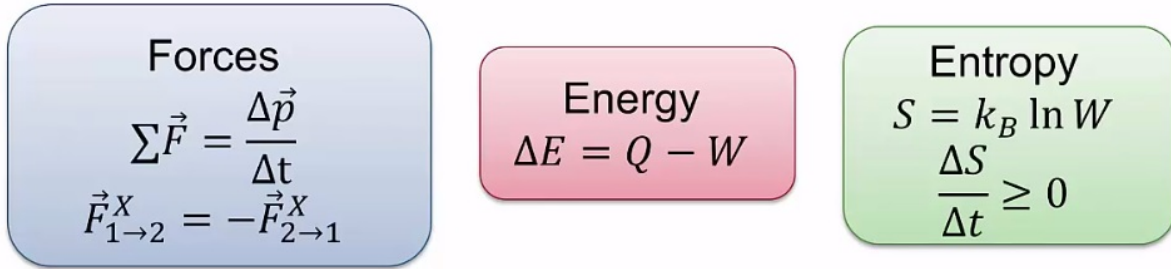
Problem: This unit overview is also available as a video [here](#).

This is our last unit of the course. Throughout this course, we have been progressing from more concrete concepts to more abstract. If you think back, we began with the concept of velocity, which people are familiar with from speedometers on their cars. Then, we move to the idea of force which people can physically feel, you can feel a push or pull, then we moved into “forces and...”, torque, impulse, work. These ideas are a little bit more abstract but rooted heavily in the idea of force. Energy is a bit more abstract of an idea, most people have some experience with energy from previous science courses, but trying to define what energy really is and getting used to thinking about it on a huge variety of distance scales can be a bit of a challenge. Now we are going to cover the one last idea of this course, which is yet still a little bit more abstract, this idea of entropy.

So, why are we covering entropy? Well, I have two answers. First, a practical answer. Many of you have seen entropy in a previous class, typically chemistry. For example, here is a slide from Chem 112 at UMass, and many of you will see this idea again. Entropy changes can help determine if reactions proceed spontaneously or not, and appears in the all-important Gibbs free energy. Typically, up to this point, you’ve either done qualitative arguments about entropy increases or decreases, or looked up standard entropies of formation in tables, but what is this quantity that you’re using, in, say, Chem 112? Well, the typical answer in many courses that introduced the idea of entropy is disorder, but what is disorder? How do you quantify it? And disordered by whose perspective? Disorder, that’s a very nebulous idea. Who gets to decide what’s an ordered state and what’s a disordered state? And it turns out that this definition isn’t even correct, so I

think that if you're going to deal with the topic, as much as many of you will deal with the idea of entropy, you should know what it is.

There's another, second more physicist answer as to why I think we should cover entropy. In a sense the whole discipline of physics, not just this course, but the whole discipline of physics, from this course to the very frontiers of modern research can really be boiled down to a few key ideas.



Forces and Newton's laws, here written as $\Delta p/\Delta t$ and Newton's third law, energy and its conservation, and the last one is entropy. So, in this green box, we have the definition of entropy here and what's known as the second law of thermodynamics. No matter how much physics you study, you're still looking at how different objects respond to forces, how energy is conserved, what is the entropy in the system and how is it changing. Since forces, energy, and entropy are three of the fundamental pillars of physics, I feel it would be remiss to leave entropy out.

So, what is entropy then? If it's not disorder what is it? Well, let's think about what is going on at the microscopic level. At the microscopic level, things are of course always changing. Molecules are moving around, chemical reactions are always proceeding, but many of these changes do not affect the microscopic picture. For example, from chemistry, when you add two reactants, the reaction never really stops, we just reach an equilibrium point where the number of reactions going in one direction equals the number of reactions going in the other direction. The molecules are constantly interacting with each other, forming bonds, dissociating bonds. At the microscopic picture, we have a hubbub of activity but at our macroscopic scale we don't see a lot of change. So, what do I mean when we say we don't see a lot of change at the macroscopic level? We mean the total energy in the system, the pressure, if it's a gas, the volume, all these

types of quantities that are easily measured in macroscopic level. Entropy is the number of ways that I can rearrange things on the microscopic level, which we call the number of microstates, which we will indicate by a letter W (no, this is not the work W , this is a different W , conventions are conventions). So, how many ways can I rearrange things, how many different microstates, are there that don't change the macroscopic world: that is what entropy is. So, it turns out that counting the number of ways energy can be distributed microscopically while leaving the macroscopic world unchanged has important implications, which is weird when you stop and think about it. I mean, the number of possible ways I can arrange things seems like a very theoretical construct, and to make matters more interesting, the numbers we'll be dealing with will be ginormous. 10^{23} is not a surprising number to deal with when you start talking about the number of ways to arrange energy amongst all the molecules in a room. These types of numbers start to appear. That is a one with a mole of zeros after it. That's a big number. These huge and seemingly theoretical numbers are the basis of what entropy is.

So, what do I want you to get out of this unit? I want you to have a beginning of a grasp of what entropy is and how we can quantify it. I want you to understand why some processes proceed spontaneously due to entropy considerations. And finally, I want you to understand how entropy can drive processes in a way that results in final states that might seem more ordered to us, but are in actuality an increase in the number of microstates when you consider the whole system. The following prep videos and reading and homework problems will lay the groundwork of some of the basic mathematics you will need to study this topic. This concludes this video.

Why Probability Matters

The topic of probability often receives little emphasis in the math classes that are part of the program for biologists -- until they are required to take a serious course in statistical methods. At that point, the focus may be on the rules and formal tools for generating a statistical result rather than on making sense of what probability and statistics are telling us. This is extremely unfortunate, since the concept of probability is fundamental in a variety of situations that are of extreme importance to many biological professionals. In this essay we discuss briefly why we are interested in probability in the context of a physics class, why researchers (of any ilk) need to know probability, and why medical professionals need to understand probability.

The basic idea of probability

The basic idea of probability is about situations that occur multiple times but have factors that we cannot control. When you flip a coin, the laws of Newtonian physics could tell you where it is going to go – and whether it will land heads or tails. That is, if you knew the initial position, upward speed, initial angular orientation, and rotational velocity to a very high accuracy. AND if you knew that the coin were a fair coin (perfectly symmetric and balanced), AND that there was no breeze, etc., etc.

Even in the context of systems that are well-described by Newtonian physics, there are many systems that we cannot predict well. Their motion is just too sensitive to factors that we cannot control.

In such a situation, what do we do? We might just give up and say: “that’s not predictable,” but another approach has developed, driven by mathematicians responding to questions from gamblers in the 17th century. (Really.) In this approach we carry out the following two steps:

Determine what results are equally probable;

Count the number of ways that a result we are looking for can be made up of the different equally probable results.

For example, consider throwing two cubical dice, each with 6 sides and the sides having 1, 2, 3, 4, 5, and 6 spots respectively. When the dice are thrown so that they bounce around in uncontrollable ways, one result comes up on each. The total will range from 2 (a one comes up on each) to 12 (a six comes up on each). But each total is not equally probable – each face of each die is assumed to be equally probable. As a result there is only one way to create the result of “2” – each die has to show one spot. But there are six ways to create a total of “7” – 1+6, 2+5, 3+4, 4+3, 5+2, and 6+1, with the first number showing the result on the first die, the second the result on the second. This means, that *if we throw the dice many times we expect to get the result 7 six times as often as the result 2.* Understanding this ratio is crucial if you are going to not lose too much money playing dice!

Note a few key ideas:

- The result given by a probabilistic law does NOT tell you what will happen in any given experiment (trial); it will only tell you if you REPEAT the experiment many times. And then it will only tell you what fraction of the time you can expect different results.
- The states that are the result of our experiment do not specify every variable. There are “hidden” uncontrolled variables that we do not specify
- The model we have of the system is crucial – what are the hidden variable states (microstates) that are equally probable, and how many different ways can a result state (macrostate) be made up from different hidden variable states.

So the very nature of the “law” we are creating is different from many of the laws we are accustomed to learning in science classes – at least in the intro classes. They only tell the result of many equivalent experiments – an *ensemble* – not of an individual one.

Probability

- Defining probability in terms of an infinite number of trials
- Calculating the mean and standard deviation for results with different amounts of probability
- Defining microstate
- Defining macrostate

Exercise:

UMASS
AMHERST Instructor's Notes

Problem:

Your Quiz will Cover

- Defining probability in terms of an infinite number of trials
- Calculating the mean and standard deviation for results with different amounts of probability
- Defining microstate
- Defining macrostate

umdborg / Probability (2013). Available at:

<http://umdborg.pbworks.com/w/page/68375351/Probability%20%282013%29>.

(Accessed: 1st August 2017)

Introducing probability

Probability is one of those words for which we all have an intuition, but which is surprisingly hard to define. For example, in our discussion of [diffusion](#), we make the assumption that particles move either to the left or right with equal probability. But try to define what this means. Try, more concretely, to define what it means for a coin to have an "equal probability" of coming up heads or tails—but without using words in your definition that are synonymous to "probability" (such as "chance" and "likely"). It's really hard to do! In fact, entire branches of philosophy have been devoted to the question of how to

define what is meant by "equal probability". So if you find yourself thinking hard about the probabilities we encounter, you are in good company!

The key idea in probability is **lack of control**. When we flip a coin, it's extremely hard to control which way the coin will come down. The result is very sensitive to the starting conditions you give it at a level of sensitivity greater than you can control. Which, of course, is the point of flipping a coin.

One definition of "equal probability" might look something like this:

As the number of tosses of a fair coin approaches infinity, the number of times that the coin will land heads and the number of times that the coin will land tails approach the same value.

Is that a useful definition? Maybe, but it doesn't seem to capture everything that we intuitively know to be true. We'd like to know what the chances are that the coin will land on "heads" when we toss it just once, without having to toss it an infinite number of times. And we all have the feeling that the answer is obvious - it's $\frac{1}{2}$! - even if we have a hard time expressing it rigorously.

How would we know if it's fair?

Of course determining whether a coin is "fair" or not would require testing it an infinite number of times. And in the real world we expect that no real coin would be perfectly fair. It might be a tiny bit unbalanced so that it consistently, over many many flips, comes out 0.1% more heads than tails. Would we accept that as a "fair" coin?

One of the interesting questions of probability is "how do you know" that a coin is fair, for example? Or better: how well do you know that a coin "appears to be fair"? This subject carries us beyond the scope of this class into the realm of [Bayesian Statistics](#). We won't discuss that here, though we will note that Bayesian analyses play a large role in the modern approach to medical diagnosis and both medical students and biological researchers will eventually have to master this subject!

A simple model for thinking about probabilities: a fair coin

Rather, we will make a simplified model that we can analyze in detail mathematically. We will assume that we have a (*mathematically*) *fair coin* -- one that if it we flipped it an infinite number of times would come up an equal number of times heads and tails.

Now we can get back to our story. Let's see if we can make some interesting observations about probabilities by relying on just our intuitions. Suppose, for example, that I toss a (mathematically) fair coin ten times. How many times will it come up "heads"? The correct answer is: who knows! In ten flips, the coin may land on heads seven times, and it may land on heads only twice. We can't predict for sure. But what we *do* know is that if it is a fair coin it is more *likely* that it will land on heads 5 times than it is that it will land on heads all 10 times.

But why do we feel that is the case? Why is the result of 5 heads and 5 tails more likely than the result of 10 heads and 0 tails? If each toss is equally likely to give heads as it is to give tails, why is the 5/5 result more likely than the 10/0 result?

The answer is that there are many more ways for us to arrive at the 5/5 result than there are ways for us to arrive at the 10/0 result. In fact, there is precisely ONE way to arrive at the 10/0 result. Note that in stating "5/5" we are assuming that we don't care in which order the heads and tails appear -- we only care about the total number.

If we only care about the totals: microstates and macrostates

If we only care about the totals there is only ONE way in which you would arrive at the result that the series of tosses produced 10 heads: HHHHHHHHHH. You have a 50% chance the first flip will be a head, a 50% chance the second will be a head, and so on. Therefore the probability of 10 heads is $1/2^{10}$ or 1 in 1024.

On the other hand, here are just a few of the 252 ways of arriving at the 5/5 result: HHHHHTTTTT, HTHTHTHTHT, TTTTTHHHHH, HTTHHTTHTH. Each of these particular strings also only has the probability of 1 in 1024 to come up, since there is a 50-50 chance of a head or a tail on each flip. But since there are 252 ways of arriving at 5/5 the chance of finding 5/5 (in any order) is

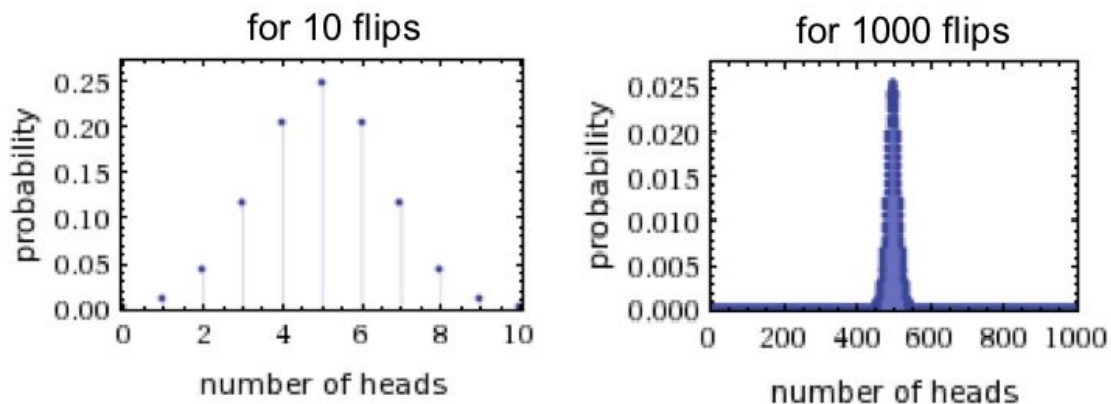
252/1024 -- much greater than finding 10/0 and in fact greater than finding any other specific mix of heads and tails.

Another way of expressing the probabilistic intuition we have been describing is to say that a system is much more likely to be in a state that is consistent with many "arrangements" of the elements comprising the state than it is to be in a state consistent with just a few "arrangements" of the elements comprising the state. An "arrangement" in the coin toss discussion corresponds to a particular ten toss result, say, HTTHTTTHHHT. The 10/0 result is consistent with only one such arrangement, while the 5/5 result is consistent with 252 such arrangements.

The difference between a specific string of heads and tails and the total count (in any order) is a model of a very important distinction we use in our development of the 2nd Law of Thermodynamics. The specific string, where every flip is identified, is called a **microstate**. The softer condition -- where we only specify the total number of heads and tails that result -- is called a **macrostate**. In our mathematical fair coin model, what "fair" means is that *every microstate has the same probability of appearing*.

What happens as the number of tosses increases from 10 to, say, 1000? As you might guess, it becomes even more likely to obtain a result near 500/500 than it is to obtain a result near 1000/0. In the jargon of statistics, the probability distribution gets "sharper."

Distribution of number of heads (assuming a fair coin)



In chemical and biological systems we often deal with HUGE numbers of particles, often on the scale of moles (one mole of molecules contains more than 10^{23} particles!) so one can imagine what the probability distribution looks like

in such cases. It is incredibly sharp. The only macrostate that we ever see is the most probable one. Regularities emerge from the probability that are (as long as we are talking about many particles) as strong as laws we consider to be "absolute" instead of "probabilistic".

Exercise:

UMASS
AMHERST Instructor's Notes

Problem: The following is also available as a video [here](#).

Probability, Means and Standard Deviations

Let's begin by giving a little bit more thought to the idea of probability that you've explored in some of your readings. The probability of an event is the fraction of time it occurs if the process is repeated an infinite number of times. For a coin, for example, if we flip it fair coin an infinite number of times we expect that one half of them will be heads. Similarly, for a dice, we expect that if we roll it an infinite number of times, one-sixth of the rolls will be a two. Colloquially, the higher the probability, the more likely an outcome is to occur.

If one event does not affect the next, then we say that the events are independent. In this course, we will only be dealing with independent mutually exclusive events. Let's begin by thinking about an example of interpreting the idea of probability. Say you roll a fair dice. What is the probability that you will roll a six? Well, of course the answer is one out of six. If you were to roll the dice an infinite number of times that, you would observe that one sixth of the rolls would in fact be a six. Now, let's say you have rolled a dice three times, and the result of each roll has been a six, i.e. you have rolled three sixes in a row. What is the probability that your next roll will also be a six? Well, the answer to this is still $1/6$ th. Each roll is independent of the previous, so your probability of the next roll being a six is still one out of six, regardless of what has happened in the past. Dice don't have memory, they don't remember, so the odds of your next roll being a six are one out of six.

Now with this idea of probability, let's move on to thinking about how to calculate means of events with differing probabilities. Consider the following

set of measurements for the height of the library, as measured, in meters:

88,87,88,90,90,88,85

We know how to calculate the average of a set of numbers; you add up all the numbers and then divide by the number of measurements. In this example, we would add up 88, 87, 88, 90, 90, 88, and 85 and divide by 7, to get an average of 88, but we see in this data set that each result appears to not be equally probable. 88 occurs three-sevenths of the time, and 90 occurs two-sevenths of the time. Well, we can deal with this as we just did by adding all the numbers up and counting 88 three times, or we can readjust our definition of average to include the idea of probability:

$$\mu = \sum_i^n p_i x_i$$

In this new definition, we don't just add up the events, we add up the probability multiplied by the value. So, we take each value multiplied by the probability, and then add to get the mean. In this example, we say that the probability of 88 is three out of seven, so we multiply 88 and 3/7. The probability of 90 is two out of seven, and so we multiply 90 by 2/7. 87 and 85 both have probabilities of one over seven, and so we multiply 87 and 85 by 1/7. If you churn this out in your calculator, you will see that you get the exact same result of 88. So, clearly these two methods yield the same result, however, the second is more powerful if we don't know the full data, but, say, only know the probabilities of different outcomes.

Now let's move on to thinking about calculating standard deviations of events with different probabilities. Here in this table,

Value	Probability
2	0.2

Value	Probability
4	0.4
6	0.1
8	0.3

What is the standard deviation of these data? Well, in our formula for mean, all we did was we change the $1/N$ to the probability of a given event. You would do the same thing for standard deviation. You do the same thing for standard deviation; instead of multiplying by $1/N$ out front, you bring it inside the sum, N multiplied by the probability. So now, this equation says take each event, subtract the mean, square it, multiply by the probability, and add them all up, and that will give you the standard deviation squared. Let's test this formula using these data. We would begin by calculating the mean itself, because the mean is an element of calculating the standard deviation. So, to calculate the mean, we say the mean is the sum of the probability of an event multiplied by the value. In this case, let's carry out this calculation for these data.

$$(2)(0.2) + (4)(0.4) + (6)(0.1) + (8)(0.3)$$

Evaluating this expression gives us a mean of 5.

So, now that we have a mean, we can proceed to calculating the standard deviation. The way I'm going to do this is I'm going to add a column to my table, $x - \mu$, for each value.

Value	Probability	$(x - \mu)$
2	0.2	-3
4	0.4	-1

Value	Probability	$(x-\mu)$
6	0.1	1
8	0.3	3

In our definition of standard deviation, we care about this value squared, so, let's continue and add yet another column, squaring, which will get rid of the negatives:

Value	Probability	$(x-\mu)$	$(x-\mu)^2$
2	0.2	-3	9
4	0.4	-1	1
6	0.1	1	1
8	0.3	3	9

Now we want to multiply each value of x minus μ squared by the probability. So, I'm going to add yet another column, probability times $(x-\mu)^2$.

Value	Probability	$(x-\mu)$	$(x-\mu)^2$	$p(x-\mu)^2$
2	0.2	-3	9	1.8
4	0.4	-1	1	0.4

Value	Probability	$(x-\mu)$	$(x-\mu)^2$	$p(x-\mu)^2$
6	0.1	1	1	0.1
8	0.3	3	9	2.7

Adding these numbers up as instructed gets me a standard deviation squared of 5; turns out that for this data set, the standard deviation squared, and the average are the same. That will not generally be true. I get the standard deviation itself by taking the square root of the standard deviation squared, giving me a standard deviation of 2.24.

In summary, the probability is the frequency something occurs after an infinite number of trials, and colloquially, we say that the higher the probability, the more likely a given event is to occur. With this idea of probability, we can adjust our definitions of mean and standard deviation by swapping out the $1/N$ out front, and instead multiplying inside the sum by the probability of each occurrence.

The Meaning of "And" and "Or" in Probability

- Calculating the probability of combinations of events

Exercise:



Problem:

Your Quiz will Cover

- Calculating the probability of combinations of events

This section is also available as a video [here](#).

“And” and “or” have important roles in probability and entropy, and we will be exploring them in this section.

Consider a bowl with four balls of different colors, red, green, orange, and blue. If we reach in and grab a ball, the probability of grabbing each color is easy to see; there’s a $1/4$ chance of grabbing red, $1/4$ chance of grabbing green, $1/4$ chance of grabbing orange, and $1/4$ chance in grabbing the blue. Regarding the probability, we can say that the probability of grabbing each color is $1/4$.

Now let’s think about the probability of grabbing a blue ball or grabbing a green ball. There’s four different possibilities when you pull out a ball, red, green, orange, and blue, and two of them are either blue or green, and so the probability becomes $2/4$. Thinking about this mathematically, you’ll notice that the final probability is the sum of the two probabilities. In this problem, we’re looking for the probability of pulling a green ball OR a blue ball, and a general guideline is that when you see “or”, that usually tells you to add.

Moving onto “and”, what is the probability of pulling out a blue ball, putting it back, and pulling out a green ball? Let’s go through all the

different possibilities if you pull the blue ball out first:

blue, red	red, red	green, red	orange, red
blue, green	red, green	green, green	orange, green
blue, blue	red, blue	green, blue	orange, blue
blue, orange	red, orange	green, orange	orange, orange

Out of all these possibilities, there is only one case where you pull out the blue ball, and then the green ball, so the probability is 1/16. Notice that this is the probability of each ball multiplied together. Just like how you add when you see “or”, you multiply when you see “and”.

Now let’s look at a more complicated example that combines these two ideas. What is the probability of pulling out the blue ball and then the green ball, or pulling out the green ball and then the blue ball? Essentially, it’s the same example as above, only we don’t care about the order in which the ball is pulled. First, let’s apply these ideas of adding with “or” and multiplying with “and”. In this scenario, we are looking for blue AND green, OR green AND blue. We can model this as:

$$(P_b * P_g) + (P_g * P_b)$$

where P is the probability. Plugging in our values of the probabilities:

$$(1/4 * 1/4) + (1/4 * 1/4)$$

Solving this gives us 2/16, or 1/8. We can confirm this by looking back at table above, we can see that there’s two possibilities out of the 16, so the probability is 2/16 here as well. In general, the “and” and “or” rules will work; there are very few cases where this will not apply.

Statistical Distributions

- Interpret the probability of different events using statistical distributions

Exercise:



Problem: Your Quiz will Cover

- Interpreting the probability of different events using statistical distributions

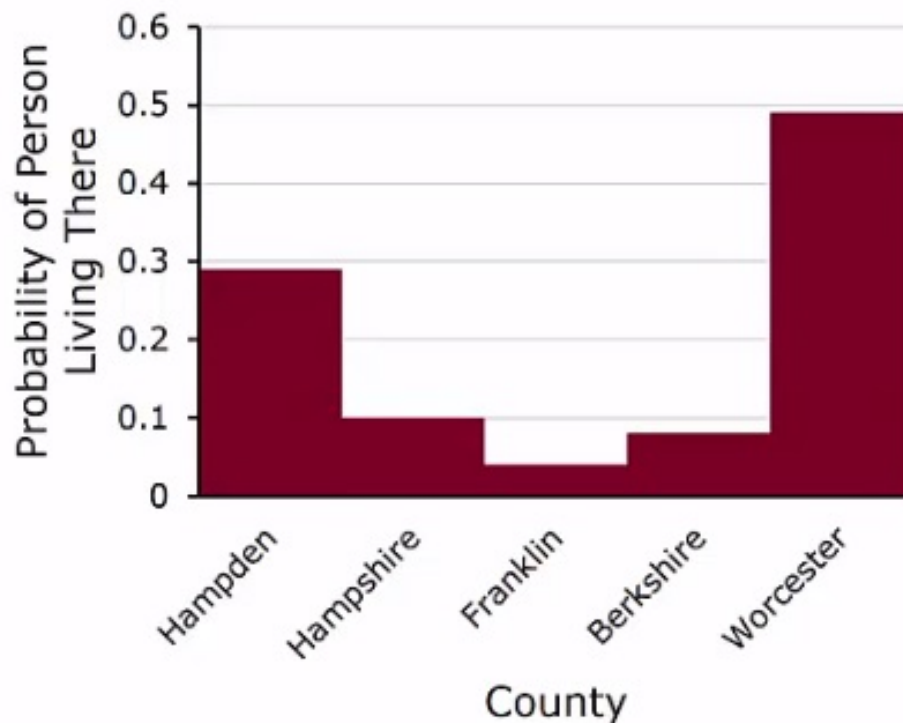
Let's begin by looking at the distinction between a discrete and a continuous variable. Discrete variables cannot take on any value. For example, coin tosses are either heads or tails; you can't be half and half. Similarly, a dice will return one of the values one, two, three, four, five, or six. A dice will never return 2.342, for example. Continuous variables, on the other hand, can take on any value within a given range, and the limit of precision is not from an intrinsic limit within the system but from the measurement technique. Some examples of continuous variables include height. It is completely possible for a person to be 156.03423 centimeters tall, and we can measure a person's height theoretically to any degree of precision that we would like.

Similarly, we often consider mass to be a continuous variable. Now, while strictly, mass is going to be discrete because you can't have less than one electron's worth, the resolution is so small that we generally consider mass to be a continuous variable.

Let's begin by thinking about, now, let's move on to thinking about probability distributions beginning with discrete data.

Western Mass County	Probability of person living there
Berkshire	0.08
Franklin	0.04
Hampden	0.29
Hampshire	0.10
Worcester	0.49

Here, I have a table of the probability of a given person living in one of the five western Massachusetts counties, Berkshire, Franklin, Hampton, Hampshire, and Worcester. A probability distribution is a bar graph where the height of the bar is the probability of the occurrence. So, for these data, a probability distribution would look something like this.

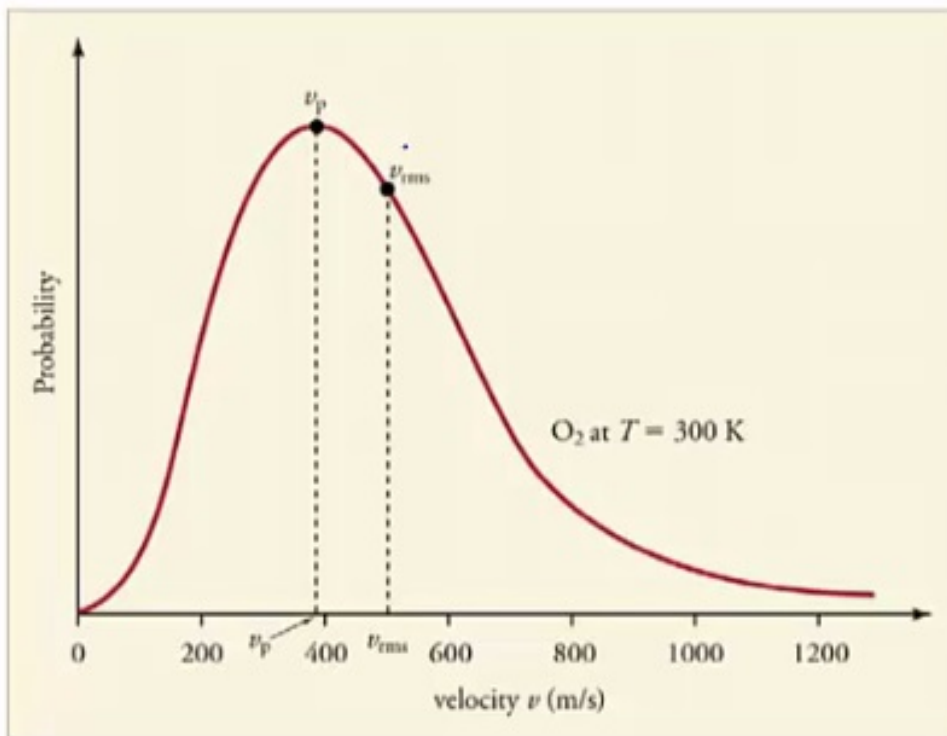


We can see each county listed on the horizontal axis, and the height of the bar indicates the probability of the person living there. All possible outcomes are in fact listed, since all possible outcomes are listed, the heights of all bars together must equal 1, because the probability of a person who lives in western mass living in one of these five counties is of course one hundred percent. Let's think about how to use these types of graphs. So, let's begin with this question as an example. What is the probability of a person who lives in Western Massachusetts to live in either Hampshire or Berkshire counties? We can read off the graph that the probability of a person living in Hampshire County is 10%, 0.10 while the probability of a person living in Berkshire County is a little bit less 0.08. The "or" tells us that we should add, in line with a previous video. So, the probability of a person living in Hampshire or Berkshire is the sum of the probability of a person living in Hampshire and the probability of a person living in Berkshire. Add these two probabilities together, and you get a probability of 0.18 or 18%. This is something that you need to be able to do.

Now let's think about probability distributions for continuous variables. Remember, continuous variables are those quantities that can take any value. An example of a continuous variable might be particle speeds at a

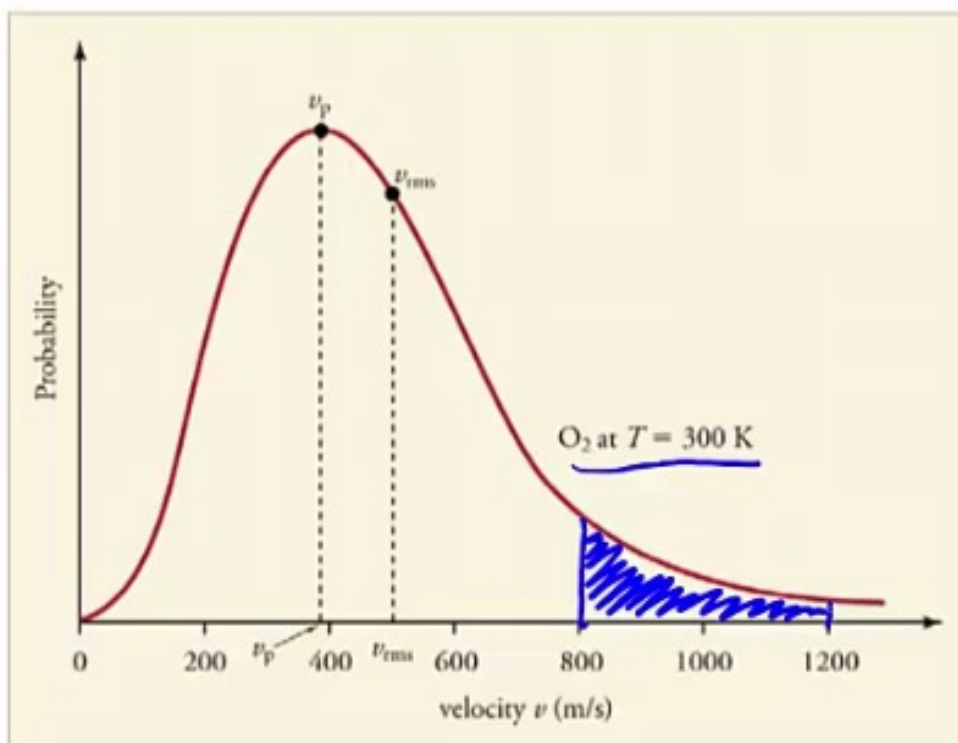
given temperature. At a given temperature, particles have a huge variety of different speeds. The expression $K = \frac{3}{2}K_bT$ tells you the average kinetic energy, or the average speed, v_{rms} . So, let's think about how to interpret probability distributions of continuous variables. The probability of any given number is 0. To understand this, think about what is the probability that a molecule bouncing around the room you're sitting in has exactly 400 meters per second worth of velocity, 400 meters per second to an infinitely high level of precision. Zero, it will always deviate from 400 by a little bit. Because of this, probability is only meaningful if we speak about range of values. Thus, to get the probability, we look at the area under the curve between the values we're interested in.

For example, what is the probability that the velocity of an oxygen molecule at 300k has a velocity between 800 and 1200 meters per second?



Well, we have the probability distribution. The area we are interested in is the area between 800 and 1200. So, this area here tells us the probability that a given oxygen molecule has a speed within this range. Since all possible speeds are represented, the area under the entire curve will be

equal to 1. The word associated with this is we say that the curve is normalized. You need to be able to recognize that probability is an area.



Factorials

- Calculating a factorial
- Approximating the value of a factorial using the Stirling Approximation

Exercise:

UMASS
AMHERST Instructor's Notes

Problem:

Your Quiz will Cover

- Calculating a factorial
- Approximating the value of a factorial using the Stirling Approximation
- $\ln(AB) = \ln(A) + \ln(B)$
- $\ln\left(\frac{A}{B}\right) = \ln(A) - \ln(B)$
- $\ln(A^N) = N \ln(A)$

This section is also available as a video [here](#).

What is a Factorial?

Let's begin by thinking about what a factorial is. In our next unit on thermodynamics and statistical physics, a lot of our time will be concerned with the study of entropy, what it is, and how can we quantify it. Over the course of this study, we will encounter a mathematical operation which is known as the factorial. The factorial comes up when you are looking at probabilities quite frequently. For example, when answering the question "if I flip a penny, a nickel, a dime, and a quarter, how many possible permutations of two heads are there?", you will need to use the idea of the factorial to solve it. The factorial operation is indicated by the !, such as 3!. How do we calculate factorials? Well, 5! just means to take 5, multiply it by

4, multiply that by 3, multiply that by 2, and finally, multiply by one giving us an answer of 120. Meanwhile, 500! means 500 times 499 times 498 times 497 and so on and so on and so on and so on, until eventually 5 times 4 times 3 times 2 times 1. Now, my calculator has a factorial button; when I try to put in 500!, it essentially burst into flames, but you can see the idea of the calculation here. You take the number, subtract 1, and multiply, and you repeat this process until you get down to 1. Because of this operation, the answer to factorials can get very big very fast.

Before we come to a method on how to deal with such large numbers, there's one last point to bring up; 0! is equal to 1. This is a convention, and we won't go into why necessarily, but you need to know that 0! is equal to 1. Factorials give us very large numbers, so when we take the factorial of a large number, we get a ginormous number, and we need to think of ways to do this without setting our calculators on fire. In our study of thermodynamics and statistical physics will be looking at molecules. Molecules come in moles, which is 10^{23} so we will be looking at 10^{23} ! Again, if you try to do this with the calculator's factorial button, you probably will get an error, and so we need a way to handle this.

The Stirling Approximation

Fortunately, in our study, we'll only be interested in taking the natural logarithms of factorials, so if we're interested in $n!$, what we'll really be interested in is the $\ln(n!)$. This will save us quite a bit, because if we're interested in the natural log of the factorial of a number, then we can use what is known as the Stirling approximation.

$$\ln(N!) \approx \frac{1}{2} \ln(2\pi) + N + \frac{1}{2} \ln(N) - N$$

This formula is on your equation sheet, so you don't have to memorize it, however, you need to be able to use it. We're going to try it through a few numbers to see how well it does. So, here is the calculation for two different values of n , 10 and 20.

N	N!	ln(N!)	Stirling Approximation	% Difference
10	3628800	15.10441	15.09608	-0.0552%
20	2.43×10^{18}	42.33562	42.33145	-0.0098%

You can see even at 10, 10! is getting very large. The natural logarithm of 10! is about 15.1. If I use the Stirling approximation, I get essentially 15.1. The difference is very tiny, as you can see by the percent difference. So, the difference between the natural logarithm of 10! and the Stirling approximation is 0.05%. Similarly, with the calculation with 20, 20! is already into 10^{18} , which gives us a natural logarithm of 42.3, and now the difference between the natural logarithm of 20! and the result of the Stirling approximation is even smaller, less than one one-hundredth of one percent, so the Stirling approximation is quite accurate.

Combinations

- Determining the number of ways to arrange n items into subsets of r size

Exercise:



Problem: Your Quiz will Cover

- Determine the number of ways to arrange n items into subsets of r size

In order to solve combinations, you need to know how to handle factorials, so make sure you understand the material in the previous section before moving on. This section is also available as a video [here](#).

What are the types of questions we are looking to answer with combinations? One example is, say that I have five apples of different varieties, how many different combinations of three apples are there? Another example question would be that there are four teams in an NFL division, how many games are necessary for each team to play every other team in their division? Or another example is, say I have 10 molecules in a box, how many different combinations are there with three molecules on the left-hand side and seven on the right-hand side?

So, what is common amongst these different situations? These problems are looking at a large pool of items and trying to choose a subset where the order of the items is not important. In the first example, I'm trying to choose three apples out of five, it doesn't matter what order I choose the apples in. In the second example, I'm choosing two teams to play each other out of the four teams in an NFL division. It doesn't really matter what

order the teams play in in this perspective, it just matters how many games do I need. And in the last example, I'm looking to choose three molecules out of ten to be on the right. It doesn't matter which three, just that there are three.

So, now let's move on and try and calculate these different combinations. I will explore this in the Apple example, wherein I have five apples of different varieties and want to know how many different combinations of three apples are there. The way to calculate the number of combinations is given by the formula

$$\frac{n!}{(n-r)!r!}$$

where n is the number of objects in total, in this case five for the five apples, and r is the number of objects at your sub group, in this case three, because I want three apples. Plugging in the numbers into this formula, we see

$$\frac{5!}{(5-3)!3!}$$

which is

$$\frac{5!}{2!3!}$$

Calculating out the factorials, we get

$$\frac{120}{2 * 6}$$

which means that there are 10 different combinations of three apples, given the five that I have.

This calculation is called a binomial coefficient, for reasons that are somewhat sophisticated, and this calculation can be represented in several different ways, and you should be familiar with all the different ways of

representing this calculation, as different fields tend to use different notation. The calculation can be represented as:

$${}_nC_r$$

nC_r

$$\binom{n}{r}$$

These all mean the same formula for calculating combinations.

Introduction

Exercise:

UMASS AMHERST Instructor's Notes

Problem:

Your Quiz will Cover

- Define entropy in terms of the number of microstates
- State the fundamental principle of statistical mechanics; that all microstates are equally likely
- Calculate the entropy for simple configurations
- From the entropy, determine the number of microstates available

These topics apply to this entire chapter, so keep these in mind as you read on.

Entropy is a weird and abstract concept. However, you might have encountered entropy before. The entropy we will discuss in this class is the same entropy in Gibbs free energy that you may have seen in one of your chemistry classes. You may have also heard of entropy before as being “disorder”. If this is the case, *please forget this idea*, as a lot of confusion surrounding entropy involves this concept of disorder. The idea of entropy as disorder is an outdated one, and is sometimes simply not true.

Consider a bowl of ice sitting in water. Eventually, the ice will melt, and we’ll be left with just a bowl of water. While a bowl of ice inside water may seem more “disorderly” than a uniform bowl of water, the water is a higher entropy state than the water with ice. We’ll discuss why this is in more detail in class.

Getting a handle on the concept of entropy will take some time. We are not expecting you to fully understand it through the readings, and we will work in class to help develop this concept. As always, what we do expect you to

know will be listed at the top of each section, under the UMass instructor's notes.

Gibbs Free Energy

Exercise:

UMASS AMHERST Instructor's Notes

Problem:

This section is also available as a video:

<https://www.youtube.com/watch?v=2WlkLcGbbfo>

The Gibbs free energy is a concept that often comes up in chemistry and biology courses. The definition of the Gibbs free energy is that the change in the Gibbs free energy

$$\Delta G = \Delta H - T\Delta S$$

Let's deconstruct this equation a minute, ΔG is the change in the Gibbs free energy, this new quantity that we're introducing, ΔH is the change in enthalpy, not the change in energy, the change in enthalpy that we discussed back in Unit 4, but you may remember from unit 4 that when we're talking about processes under constant pressure, and most everything is at constant pressure in biology, as most things are open to the air and therefore take place at air pressure, under this situation, enthalpy is the same as heat, so $\Delta H = Q$ at constant pressure, which will be the case for most processes that you will encounter and all of the processes that we will discuss.

T in this expression is what you might expect; it's the temperature. Why do we need the temperature in this equation? Well, is a heat, or enthalpy change, of five joules per mole big or small? Well, you can't say that unless you have something to compare that five joules per mole to. So, what energy in the system can you compare that five joules per mole to? Well, you could compare it to the average energy available, and as we saw in unit 4 the average energy is the temperature, so that's why the temperature is in this expression. It provides a comparison point for the enthalpy. The last quantity is ΔS , the change in the entropy. It turns out that the sign, but interestingly not the magnitude of ΔG tells you if a process will happen

spontaneously. If ΔG is negative, then the process will occur spontaneously. So, clearly if you want to use this Gibbs free energy to understand if processes will be spontaneous or not, you need to know what entropy is.

How can I get a spontaneous process using this definition of ΔG ? Well, there's two sorts of ways. One, I can be energy-releasing. If I release energy as heat, i.e. I'm an exothermic reaction, then the Q will be negative, and thus if I'm at constant pressure the enthalpy will be negative, and this will push us towards a negative Gibbs free energy change in negative ΔG . The other way is to increase the entropy. If the entropy ΔS increases, then because of this negative sign, the Gibbs free energy change ΔG will actually be pushed negative, so an increase in entropy, i.e. the number of ways to arrange things. will push towards a negative change in Gibbs free energy.

In summary, the Gibbs free energy is an important concept in the chemical and life sciences, at constant pressure which most processes in the chemical and life sciences are, the sign of the Gibbs free energy indicates if a process will be spontaneous or not. If the change in the Gibbs free energy, ΔG , is negative, less than zero, then the process will be spontaneous. The Gibbs free energy looks at a balance of the change in enthalpy ΔH and the average energy in a system, the temperature, and a change in the entropy.

Thus, entropy is an important concept for us to understand to help us understand this idea of the Gibbs free energy. We'll see along the way that this idea of understanding what entropy is will also help us gain a greater insight into certain processes.

Second Law of Thermodynamics

Note: The following is based off of umdberg / The 2nd Law of Thermodynamics: A Probabilistic Law (2013). Available at: <http://umdberg.pbworks.com/w/page/68405604/The%202nd%20Law%20of%20Thermodynamics%3A%20%20A%20Probabilistic%20Law%20%282013%29>. (Accessed: 20th July 2017)

Energy conservation -- [the 1st law of thermodynamics](#) -- suffices to rule out a lot of thermal sort of things that don't happen -- like things getting warmer without any source of warmth. But there are a lot of thermal things that don't happen that are perfectly consistent with the 1st law; like thermal energy flowing from a cold object to a hotter object. In order to codify and elaborate our understanding of these results, we turn to the ideas of [probability](#) to understand *how energy tends to be distributed*.

A probabilistic law

That seems a bit strange. What does a discussion about probabilities have to do with a physical law? Physical laws are always true, aren't they? And isn't probability really about things that are only sometimes true? Well, in many ways, molecules in physics are like multi-sided dice, and the likelihood that a particle will be located in a particular location in space (or have a particular energy) is analogous to the likelihood that a multi-sided die will land on a particular side. There are many different ways for the molecules to move, and the details of why they move in one way or another is very sensitive to exactly where they are and how they are moving -- and is very much out of our control.

The likelihood that all the smoke particles in a smoke-filled room will move as a result of their chaotic motions into one corner of the room is analogous to the likelihood that nearly all the coins in a set of 10^{23} tosses will land on

heads. It's very, VERY unlikely! If you tossed that many coins over and over again for the lifetime of the universe (14 billion years) the odds that you would see all heads is still minuscule -- totally ignorable. This extremely low probability is what transforms a "probability statement" into a "physical law."

The reason you will never see the smoke particles accumulate in one small corner is that there are many, many more ways for the smoke particles to distribute themselves uniformly throughout the room than there are ways for the particles to all be located in just one corner of the room. That said, just as it is not *impossible* for all 10^{23} tosses to land on heads, it is not *impossible* that all the smoke particles will spontaneously move to one corner of the room... just don't hold your breath waiting for it to happen.

Microstates and macrostates

More generally we can say that when the number of atoms or molecules in a system is large, the system will most likely move toward a thermodynamic state for which there are many possible microscopic "arrangements" of the energy. (And they will be very unlikely to move toward a thermodynamic state for which there are very few possible microscopic arrangements.) If this seems mysterious, go back to the discussion of [coin tosses](#) - it's a pretty good analogy. The H/T ratio (say, 5/5) -- which we refer to as a **macrostate** of the system -- is analogous to a thermodynamic state of a system, where only the pressure, temperature, and density of the molecules are specified. The different ways in which that H/T ratio can be obtained (say, HTTHTTTHHHT) -- which we refer to as a **microstate** -- is analogous to the specification of the spatial and energetic arrangement of each of the atoms/molecules that compose a particular thermodynamic state. As we saw in the coin toss discussion, if one only looks at the macrostate description, one is much more likely to get a H/T result that corresponds to a large number of arrangements. Likewise, one is much more likely to get an atom/molecule distribution that corresponds to a large number of arrangements.

The second law

The Second Law of Thermodynamics can now be stated in this qualitative way:

When a system is composed of a large number of particles, the system is exceedingly likely to spontaneously move toward the thermodynamic (macro)state that correspond to the largest possible number of particle arrangements (microstates).

There are a few really important words to make note of in this definition. First, the system must have a LARGE number of particles. If the system has just a few particles, it is not exceedingly likely that the particles will be in one state rather than another. Only when the number of particles is large do the statistics become overwhelming. If one tosses a coin just twice, there is a reasonable chance (namely, 25%) that one will obtain all heads. Secondly, the system is EXCEEDINGLY LIKELY, but not guaranteed, to move toward a state for which there are the most particle arrangements. The larger the number of particles, the more likely it is, but it is never a guarantee. Thirdly, this law does not specify the specific nature of these "arrangements." It may be that we are only interested in spatial location, in which case an arrangement corresponds to the spatial location of each particle in the system. More arrangements would then correspond to more ways of positioning the particles in space. In other contexts we may be interested in energy, and arrangements would then correspond to the set of energies corresponding to the system's constituents. In either case, the most likely thermodynamic state is the one for which there are the most microscopic arrangements.

Biological implications

The Second Law of Thermodynamics is a statistical law of large numbers. But we have to be careful. Although biological systems almost always consist of a huge number of atoms and molecules, in some critical cases there are a *very small* number of molecules that make a big difference. For example, a cell may contain only a single copy of a DNA molecule containing a particular gene. Yet that single molecule may be critical to the production of protein molecules that are critical to the survival of the cell. For some processes a small number of molecules in a cell (fewer than 10!)

can make a big difference. On the other hand, a cubic micron of a fluid in an organism typically contains on the order of 10^{14} molecules! The second law of thermodynamics is a law that is indispensable in analyzing biological systems in countless contexts; but it is essential to understand it well -- not to just use it mindlessly.

[Just as our probability that the number of Heads we got in flipping coins got narrower as the the number of flips got larger, the probability that our results are those predicted by statistical mechanics (most probable macrostates) gets sharper and sharper. The variation around that perfect probability (corresponding to an infinite number of flips or particles) is called *fluctuations*. The scale of fluctuations can be estimated crudely as about $1/(\text{square root of the number})$. So for 10^{14} molecules, our corrections due to fluctuations are about 1 part in 10^7 . Whereas, if we only have $100 = 10^2$ molecules, our fluctuations are expect to be about 1 part in 10^1 or 10%.]

Entropy

Since the number of microstates corresponding to a particular macrostate plays a critical role, we need a way to count them in order to quantify what's going on with the probabilities. The number of arrangements is so large, that it turns out to be convenient to work with a smaller number -- the [log](#) of the number of microstates. This is just like counting the powers of 10 in a large number rather than writing out all the zeros. For a very large N , the number 10^N is considerably larger! And it turns out that working with the log of the number of microstates is very much more convenient. Essentially what is happening is that when you put two systems together (imagine combining two boxes of gases into one) the number of microstates of the combination is basically the *product* of the number of microstates in each. (If we flip a coin 10 times, the number of microstates is 2^{10} . If we flip it another 10 times, the new number of microstates is 2^{20} -- the product of 2^{10} with 2^{10} .) If we take the log of the number of microstates, when we add two systems together, the logs of their number of microstates add to get the total number. This turns out to be both easier to work with and to lead to a number of nice ways of expressing things mathematically.

The log of the number of distributions of the energy that correspond to the thermodynamic state of a system is termed the "entropy" of the system, and is given the symbol S . Another way of stating the Second Law, therefore, is to say that systems are exceedingly likely to spontaneously move toward the state having the highest entropy S . Using the symbol W to represent the number of arrangements of the energy that correspond to a particular thermodynamics state, we can write an expression for entropy as follows:

$$S = k_B \ln W$$

The constant k_B is called Boltzmann's constant, and its value is 1.38×10^{-23} J/K. (Yes, it's the same constant we ran into in our discussion of kinetic theory of gases -- the gas constant R divided by Avogadro's number, N_A .) The important thing to take from this equation is that the entropy S is a measure of the number of arrangements W . As W goes, so goes the entropy.

But of course the number W is usually a HUGE number, and counting up arrangements to arrive at its value would usually take you forever. Fortunately, it is very rarely the case that we actually need to do the counting. Rather, we usually need only to compare two thermodynamic states and to decide which one is consistent with the greatest number of microscopic arrangements. That is the state to which the system will evolve.

Systems

When discussing the Second Law of Thermodynamics, it is crucial to be very careful about defining the system that one is considering. While it is always the case that the entropy of the *universe* is overwhelmingly likely to increase in any spontaneous process, it is not necessarily the case that a particular sub-system of the universe will experience an increase in entropy. If the system being studied is isolated, i.e., if no matter or energy is allowed to enter or leave the system, then the system's entropy will increase in any spontaneous process. But, if the system is NOT isolated, it is entirely possible its entropy will decrease. Stated more generally, it is entirely possible that one *part* of the universe will exhibit an entropy

decrease during a spontaneous process while the rest of the universe exhibits a *larger increase* in entropy, such that the overall entropy in the universe has increased. All of this is just to say that it is of utmost importance to be clear about the system to which the Second Law of Thermodynamics is being applied.

It is not obvious at this stage that the statement of the Second Law of Thermodynamics presented here will be practically useful in understanding which processes in nature are spontaneous and which ones are not. What, for example, does any of this have to do with the fact that heat spontaneously transfers from hot objects to cold objects and not the other way around? What does this have to do with chairs sliding across a room? And what does it have to do with the electrostatic potential across a biological membrane? As it turns out, the Second Law of Thermodynamics as defined above can in fact explain those examples.

Example: Arranging energy and entropy

Note: The following is based off of

1. umdberg / Example: Arranging energy and entropy. Available at:

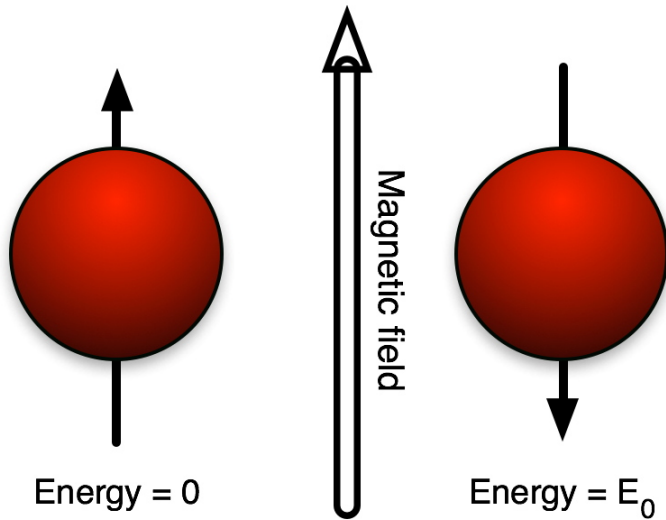
<http://umdberg.pbworks.com/w/page/104869513/Example%3A%20Arranging%20energy%20and%20entropy>. (Accessed: 20th July 2017)

The 2nd law of thermodynamics says that energy will tend to spontaneously distribute itself so that it is, on the average (and this phrase is very important -- see [How Energy is distributed: Fluctuations](#)), spread equally to all degrees of freedom. It is not easy to see what this means, so let's consider a problem and work out a simple example in detail.

The entropy of a particular macrostate is proportional to the logarithm of the number of microstates corresponding to that macrostate. To see what that means and why entropy tells us about how a system will spontaneously tend to redistribute its energy, let's consider a "toy model" -- one that is sufficiently simplified that we can understand clearly the mechanism behind the mathematics.

One of the reasons that it is difficult to understand entropy as about energy distribution is that many of the degrees of freedom we deal with -- kinetic energy in three directions, energy of rotation, ... -- are continuous. The energy in them can take any value. This makes it hard to see that entropy is actually about *counting* -- counting the number of ways energy can be distributed. That math showing this involves breaking the continuum up into bits, counting the arrangements of those bits, and then taking a limit as the size of the bits go to zero. This involves more math than we would like to get into at this point.

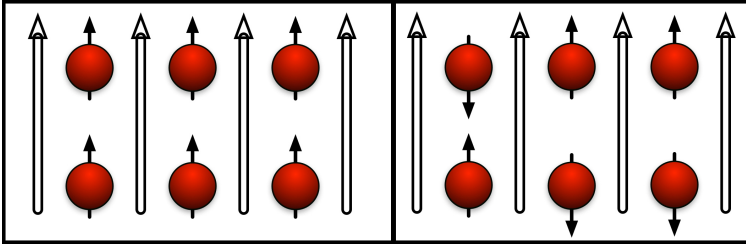
Fortunately, some degrees of freedom are not continuous: they are discrete - - their energies can only take on specific values. Here are two.



The alignment of the magnetic moment of protons in a magnetic field -- The protons that make up the nucleus of hydrogen atoms are little bar magnets.

Because of the laws of quantum mechanics, if they are placed in a magnetic field they can only either line up with or against the magnetic field. If it's lined up with the magnetic field it has a lower energy, which, if we are only discussing magnetic energy, we can call 0. (This is the way it "wants" to be.) If it is lined up in the opposite direction from the magnetic field, it has a higher energy, which we will call E_0 .*

If you start with a bunch of protons in a magnetic field and you have some energy, you can distribute it by flipping some number of protons against the field. In the figure at the left, we have 6 protons, all aligned with the magnetic field, so the (magnetic) energy of the system is 0. In the figure at the right, we have flipped 3 of the protons to be anti-aligned with the magnetic field, so the energy of the system is $3E_0$.



The orientation of the proton in a magnetic field is not only a discrete degree of freedom, it can only hold one "packet" of energy. It either has the energy E_0 or it has none. It can't hold 2 or 3 packets. It's like an on-or-off switch.

This system has relevance to how an MRI works.

Now that we have (we hope!) convinced you that a model with discrete packets of energy is "not just a toy" but also useful in real physical situations, let's solve a typical problem.

1. Consider a system consisting of four discrete degrees of freedom, each of which can only hold 1 packet of energy. Suppose we have 2 packets of energy to distribute. This system of 4 bins with 2 packets is a macrostate -- it's a system with a given amount of energy. How many microstates -- states corresponding to specific ways of distributing those energy packets -- correspond to that macrostate? What is the entropy of the macrostate with 4 bins and 2 energy packets?

Questions



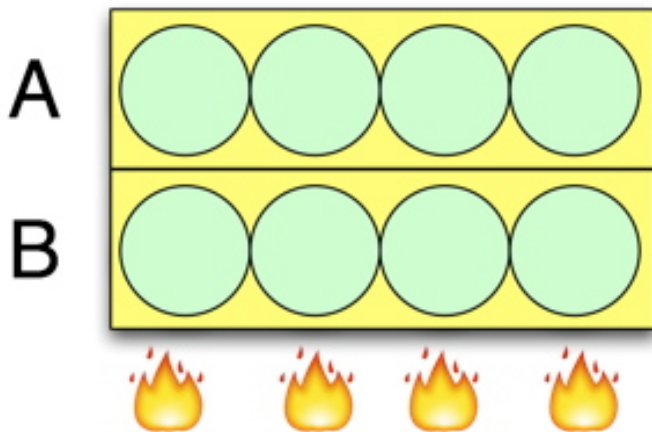
= packet of energy



= degree of freedom
(place to put energy)



2. What if we have two adjacent identical systems, A and B, of 4 bins each. What is the entropy of the state with 4 packets of energy all residing in A? Compare that to the entropy of the state with 2 packets of energy in A and 2 packets in B.



Solutions

1. How many ways can we put 2 packets of energy into 4 bins? Let's label the bins 1, 2, 3, and 4. We can only have one packet in a bin or none, so counting is pretty straightforward. We can put the first packet into our bins in any of 4 ways. The second packet can't go in the bin that we have used, so we have only three places to put it. So we have 12 ($= 4 \times 3$) ways in which we can place our 2 bits of energy into 4 places. But energy is energy. If we have one packet in bin 1 and one packet in bin 3, it doesn't matter whether we put a packet into 1 first and 3 second or the other way round. Counting 4×3 counts both those ways separately. So we have calculated each arrangement twice. The real results is half of 4×3 or 6. We can easily enumerate them: We can have the bins occupied by an energy packet as: (12), (13), (14), (23), (24), and (34). A total of 6, just as we calculated.

So there are 6 ways (microstates) of making a state with 2 energy packets in 4 bins (a macrostate). Since the entropy is

$$S = k_B \ln W$$

where W is the number of microstates, we have $S = \ln 6 k_B = 1.79 k_B$. Note that entropy has the units of k_B -- energy per degree Kelvin.

2. For the second situation, our macrostate is specified by how many energy packets are in A and how many in B. (This is like specifying the temperature of each object.) Our number of microstates is the number of ways to get that result.

For the first situation, all 4 packets in A, there is only one way we can do it. Since each bin can only hold one packet, we have to put one in each. Since there is only one way, the entropy of this macrostate is $S = k_B \ln W = k_B \ln 1 = 0$.

For the second situation, since the order doesn't matter (all energy packets being equivalent), we can put 2 packets into A first and then 2 packets into B. In part 1 we calculated that there were 6 ways to put 2 packets into 4 bins. So there are 6 ways to put 2 packets into A and 6 ways to put 2 packets into B. You can enumerate them just as we did in part 1. So how many total ways are there? Just the product of $6 \times 6 = 36$. We can easily see this by looking at what a particular microstate looks like. For A we have a

list of 6 possibilities: (12), (13), (14), (23), (24), (34). For B we have a similar list. Any AB microstate is a specification something like:

A(13)B(24). Clearly there are 6×6 possibilities. So the entropy of this macrostate is $S = k_B \ln W = k_B \ln 36 = 3.58 k_B$. (It should be no surprise that this is twice the entropy of the state found in 1.)

This tells us that the transition from all the energy in A (A hot, B cold -- entropy = 0) can spontaneously go to a state where A and B have equal energies (same temperature -- entropy = $3.58 k_B$.)

* This has the value μB , where μ is the magnetic moment of the proton and B is the strength of the magnetic field.

** The reduced mass of two masses, m_1 and m_2 , is equal to $m_1 m_2 / (m_1 + m_2)$. This adjusts the KE for the fact that both masses are moving in coordinated ways. See [Center of Mass](#) and [Diatomic Vibrations](#).

Joe Redish 2/4/16

A Way to Think about Entropy - Sharing

Note: The following is based off of umdberg / A way to think about entropy -- sharing. Available at: <http://umdberg.pbworks.com/w/page/50323410/A%20way%20to%20think%20about%20entropy%20--%20sharing>.

We've now read a lot about the second law of thermodynamics and the idea of *entropy*. The basic idea was that we are looking coarsely at a system that has a fine-grained structure that is changing rapidly in a random way. Specifically, we are looking at things *macroscopically* -- and by this we mean at a level at which the structure of matter in terms of molecules and atoms can't be seen. We only care about average properties of the molecules; things like temperature, pressure, and concentration. It is generally useful to ignore the fact that the molecules are actually moving around chaotically, colliding with each other, and chemical reactions are happening (and unhappening).

What we *are* interested in, is the following:

If two parts of a system we are considering are NOT in thermodynamic equilibrium, what will naturally tend to happen?

This is the question that the second law of thermodynamics gives the answer to. It tells us that if one part of a system is hotter than another, the natural spontaneous tendency of the system is for the temperature to even out. If a chemical reaction can occur, the reaction will continue in one direction until the rate of the reverse reaction is equal to the rate of the reaction. When the rates of forward and reverse reactions are equal, the amount of each chemical stays the same and it is called chemical equilibrium.

To understand these situations in general and to figure out which way things will happen under what conditions, we introduced the concept of entropy

and the second law of thermodynamics. The core idea is that to each coarse-grained view of a particular system (its pressure, temperature, chemical concentrations, etc. -- its *macrostate*) there are many, many different possible possible arrangements and motions of the individual molecules (its *microstate*). The idea of the second law is:

A system that is not in thermodynamic equilibrium will spontaneously go towards the state with the largest number of microstates.

The reason for this is that we assume that as the system goes through its various chaotic states, each microstate is equally probable. Therefore, the system will most often wind up in the macrostate that corresponds to the largest number of microstates.

Since the *entropy* is defined as (a constant times) the logarithm of the number of microstates, $S = k_B \ln(W)$, (see [Why entropy is logarithmic](#)), the second law can be restated as

Systems that are not in thermodynamic equilibrium will spontaneously transform so as to increase the entropy.

Well. This is an impressive sounding statement. But what does it mean? It's pretty plausible to think about flipping coins and deciding whether 5 heads and 5 tails is more or less likely to happen than tossing 10 heads in a row. But how does counting microstates help us see that hot and cold objects placed together will tend to go to a common temperature? You can do it, but it takes a LOT of heavy mathematical lifting -- and doesn't particularly help us conceptually. Another way to think about it that might help, is to think of entropy as a measure of *sharing*.

If energy is uniformly spread, it's useless.

Thermodynamic equilibrium means that the energy in a system is uniformly spread among all the degrees of freedom (i.e. distributed evenly among all places energy can go, for example, for each molecule among both its potential and kinetic energies). In such a state, the energy no longer "flows" from one set of molecules to another or from one kind of energy to another.

Thus in thermodynamic equilibrium the energy in a system is useless; no work, either physical or chemical can be done. If we want to think about how useful some energy is, we need to know not just how much energy we have, but how it is distributed. The further from equilibrium it is, the more useful it will be. We are working towards developing the idea of not just energy, but *free energy* -- useful energy.

In some sense, entropy is a measure of how uniformly the energy is distributed in a system. If the system is fully at thermodynamic equilibrium the entropy is a maximum. If the entropy is lower than than maximum, then there is room for the entropy to go up as the system moves towards thermodynamic equilibrium. The system will spontaneously and naturally be redistributing its energy toward the equilibrium state. During such a redistribution, work can get done and an organism can make a living.

Joe Redish 1/29/12

Why Entropy is Logarithmic

Note: The following is based off of umdberg / Why entropy is logarithmic. Available at: <http://umdberg.pbworks.com/w/page/49691685/Why%20entropy%20is%20logarithmic>. (Accessed: 20th July 2017)

We defined the entropy (S) of a system as $k_B \ln W$, where W is the number of possible arrangements of the system. But why? Why not just say that entropy **is** the number of arrangements? Let's think through why it has to be defined this way.

We want to define entropy as an **extensive property**, i.e. if I have two systems A and B, the total entropy should be the entropy of A plus the entropy of B. This is like mass ($2 \text{ kg} + 2 \text{ kg} = 4 \text{ kg}$), and not an **intensive property** like temperature (if you combine two systems that are each at 300 K, you have a system at 300 K, **not** at 600 K!).

What happens to the number of possible arrangements when you combine two systems? If system A can be in 3 different arrangements and system B can be in 5 different arrangements, then there are $3 \times 5 = 15$ possible combinations. They multiply! This '80s [music video](#) explains why.

So we can't just define entropy as the number of possible arrangements, because we need the entropy to **add**, not **multiply**, when we combine two systems.

How do you turn multiplication into addition? Just take the logarithm. $3 \times 5 = 15$, but $\ln 3 + \ln 5 = \ln 15$.

So that's why entropy is defined as a constant times $\ln W$. W (the number of arrangements) is a dimensionless number, so $\ln W$ is too.

The constant out in front could be any constant, but we use Boltzmann's constant, 1.38×10^{-23} J/K. When we get to Gibbs free energy, we'll see that this constant has the right units, since we need entropy to be in units of energy/temperature.

Ben Dreyfus 1/9/2012